DOCUMENT RESUME

ED 109 186 TM 004 636

AUTHOR Epstein, Kenneth I.
TITLE Sequential Plans and Formative Evaluation.
PUB DATE [Apr 75]
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D. C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Audiovisual Instruction; *Decision Making; Educational Alternatives; *Formative Evaluation; *Instructional Improvement; *Instructional Materials; Objectives; *Sequential Approach; Testing; Test Results; Validity

ABSTRACT

Sequential plans are suggested as a basis for developing decision rules for accepting or rejecting instructional materials. The technique allows inferences to be drawn concerning the effectiveness of the instruction for the target population, from formative evaluation sample data. The procedure calls for explicit statements of the required level of instructional effectiveness and the amount of error that can be tolerated in decision making. When applied to audio-visual instruction developed for the U. S. Army, sequential plans tended to require fewer students and provided a much clearer framework for making decisions than did the traditional 80/80 rule. (Author)

ED 109186

19.14

Sequential Plans and Formative Evaluation

Kenneth I. Epstein
Florida State University*

A paper presented at the American Educational Research Association annual meeting
Washington, D.C. April, 1975

* Now at the U.S. Army Research Institute for the Behavioral and Social Sciences
Arlington, Virginia

Sequential Plans and Formative Evaluation

Kenneth I. Epstein

The systematic design of instruction requires that data indicating the degree of effectiveness of instructional materials be collected and used to make improvements in the instructional materials where necessary. This process is often termed formative evaluation. Baker and Alkin (1973), in their review of the state of the art in formative evaluation, focus primarily on the problems of what types of data are useful and what one does when a decision has been made to revise the instruction. The concern of this paper is with the decision making process itself.

Baker and Alkin (1973) suggest that one of the critical factors in judging instructional effectiveness is the extent to which learners master the objectives. We can take this one step further and conceptualize instructional effectiveness in terms of the extent to which <u>any</u> student in the target population is likely to master the objectives, given the opportunity. This is exactly what is implied in the 80/80 criterion often applied to instructional development efforts. 80/80 implies that the instruction will be considered effective if at least 80% of the students who begin the instruction complete at least 80% of the objectives on the first try.

Let us assume, for the moment, that the 80/80 criterion is reasonable for a particular instructional development effort. That is, we realize that all of the students who begin the instruction may not successfully complete 80% of the objectives on the first try, but that we will be satisfied if 80% of

them do. Let us also assume that valid and reliable measures of the objectives are available. How do we determine if the instruction is acceptable in its present state? The obvious answer is to find some people who are members of the population for which the instruction is intended, let them try the instructional materials, and find out how well they perform on a test of the objectives. This obvious answer has at least one difficult question and one serious potential inadequacy associated with it. The difficult question is: "How many people do we need in the tryout group?" The potential inadequacy lies in the second half of the 80/80 criterion, 80% of the objectives. We will deal with the second problem first.

Table 1 contains the results of a hypothetical tryout of instructional material that technically meets the 80/80 criterion. Five students are tested on each of the five objectives of the instruction. A "1" signifies that the test was passed, a "0" that it was failed.

		OBJECTIVE						
		1	2	3	4	5	Total	%
STUDENT	1	1	1	1	1	0	4	80
	2	1	1	1	1	0	4	80
	3	0	1	1	1	1	4	80
	4	1	1	1	1	0	4	80
	5	1	1	1	1	0	4	80
	Total	4	5	5	5	1		
	%	80	100	100	100	20		

Table 1: Hypothetical Tryout Data

Looking at these results, we find that all five students completed at least 4 out of 5, or 80%, of the objectives. Thus it seems that we have exceeded the 80/80 criterion, and, in fact, have achieved a level of 100/80. However, these encouraging results have been obtained at the expense of objective 5. Only one of the five students completed objective 5. Can this possibly be considered effective instruction? I maintain that it cannot. In formulating decision making procedures it is potentially illogical to interpret criteria such as 80% of the students achieving 80% of the objectives literally. <u>Each objective must be evaluated independently</u>. The data in Table 1 are not indicative of acceptable instruction. The material pertaining to objective 5 must be revised.

Treating objectives independently is not only more logical than attempting to collapse over objectives, it also allows for greater flexibility. For example, suppose that all objectives were not equally critical. One could then attach different criterion levels to each objective depending on its importance. Instruction for absolutely essential skills might require that 95% or even 100% of the students accomplish the objective, while "nice to know" or interesting information might have relatively low criterion levels.

Considering the need to treat objectives independently, the 80/80 criterion rule as originally suggested must be revised. Retaining the requirement that the instruction be effective for at least 80% of the students, the new criterion rule simply states that 80% of the students must achieve each objective. This leads back to the question of the size of the tryout sample.

Suppose that instruction which is effective for at least 80% of the students in the target population is desired. The data gathered during a tryout of the instruction designed to meet this criterion is used to draw inferences about the effectiveness of the instruction for the total target

population. Clearly, the larger the sample of students in the tryout group, the better the estimate of instructional effectiveness for the total group will be. An indication of the precision with which population parameters are estimated by sample data is given by confidence limits. For example, in the case of one objective, assume that the tryout sample consists of five students, four of whom accomplish the objective. The effectiveness of the instruction, in terms of the proportion of students who accomplished the objective is 80%. However, the 95% confidence limits for a proportion based on 4 correct answers in 5 trials are 0.343 and 0.990. (These confidence limits assume that a random sample is drawn from an infinitely large population. Since most instructional development efforts involve materials which will be useful for a large number of students and since students for a tryout should be randomly sampled from the target population, this assumption seems reasonable.) The relatively widely separated values of the confidence limits imply that we should be extremely cautious in drawing any inferences about the instructional effectiveness for the total population from a tryout sample of five. Unfortunately, increasing the sample size while staying within the bounds of practical constraints doesn't help much. For example, the 95% confidence limits for proportion for observing 8 correct in 10 trials are 0.397 and 0.963; for 16 correct in 20 trials, 0.589 and 0.929; for 24 correct in 30 trials, 0.636 and 0.909; for 40 correct in 50 trials, approximately 0.67 and 0.90; and for 80 correct in 100 trials, approximately 0.71 and 0.88.

Sequential testing is an alternative to tryout situations where the sample size is fixed before the tryout begins, and no decision is possible until all the data have been analyzed. The use of a sequential testing

strategy takes advantage of the fact that a very good product or a very poor product can be expected to reveal its character when only a small sample is tested, and that more extensive sampling is only necessary for products of borderline quality. In general, the sequential testing strategy calls for observing one sample item at a time with the possibility of a decision about the total population after each observation. That is, sample items are drawn randomly, one at a time, from the population, and, based on observations of that sample, the total population is 1) accepted, 2) rejected, or 3) no decision is made. For the case where an acceptance or rejection decision cannot be made, another sample item is selected and the decision rule applied again. The process continues until an acceptance or rejection decision can be reached

In terms of formative evaluation, the total population consists of all students for whom the instruction is intended. One sample item corresponds to one student chosen at random from the target population. The performance of each student after completing the instruction is used to predict how well any other student in the target population would do, were he exposed to the instruction.

Under the 80% criterion rule the instruction will be considered effective if at least 80% of the students in the target population accomplish the objective. This criterion may also be interpreted as the probability that any randomly chosen student will accomplish the objective, that is, 0.80. In other words, the performance of a randomly chosen student may be considered a Bernoulli variable with the probability of success being equal to instructional effectiveness. An acceptance decision or a rejection decision can be made when sufficient evidence to draw inferences about the instructional effectiveness has been gathered.

Four parameters are required to develop a sequential testing plan. Two of the parameters are related to the required level of instructional effectiveness They will be designated p_1 and p_2. The other two parameters, α and β, are related to the amount of error in decision making that can be tolerated.

α is determined by answering the following question: "What percent of high quality instruction can be erroneously rejected?" β is determined by answering a similar question: "What percent of low quality instruction can be erroneously accepted?"

Notice that the model is based on the fact that absolute accuracy in decision making can never be achieved when decisions are based on sample data. The strength of the model lies in the explicit statement of the allowable error. However, the possibility of error also implies that an area of indecision exists. This means that we can no longer demand point estimates of instructional effectiveness. Rather, we must define what is meant by unacceptably low quality and unquestionably high quality. The area in between low quality and high quality is, in effect, an area of indifference or indecision. For example, if we want to restate the 80% criterion in terms useful to sequential testing, we might say that instruction which is effective for 90% of the target population is of unquestionably high quality but that instruction which is effective for only 70% of the target population is certainly unacceptable. Instruction which is effective for between 70% and 90% of the target population we are indifferent about. It may or may not be acceptable. The parameters p_1 and p_2 specify explicitly what is meant by "high quality" and "low quality." By varying the values of p_1 and p_2, α and β it is possible to specify in great detail exactly what is required

of an instructional program. The model is also readily adaptable to more elaborate systems incorporating differential loss functions.

The original development of the mathematics for sequential testing strategies was described by A. Wald in his book Sequential Analysis (John Wiley & Son, 1947). The equations which follow were adapted from the discussion of sequential testing in Lindgren and McElrath (1966) and Crow, Davis and Maxfield (1960).

The sequential likelihood ratio test is designed for testing between two simple hypotheses which, in the case of a Bernoulli population, can be written:

$$H_1 : p = p_1$$

$$H_2 : p = p_2$$

where p_1 = 1- the effectiveness of high quality instruction and p_2 = 1- the effectiveness of low quality instruction. In other words, p_1 represents the probability that a student will be unable to accomplish the objective when the instruction is of unquestionably high quality, and p_2 represents the probability that a student will be unable to accomplish the objective when the instruction is of unacceptably low quality.

The test is based on the value of the likelihood ratio computed after each observation, including all of the observations obtained up to that point:

$$(1) \quad \lambda_n = \frac{L(p_2)}{L(p_1)} = \frac{p_2^k (1-p_2)^{n-k}}{p_1^k (1-p_1)^{n-k}}$$

where λn = the likelihood ratio

$L(p_2) = p_2^k (1-p_2)^{n-k}$ = the likelihood of p_2

$L(p_1) = p_1^k (1-p_1)^{n-k}$ = the likelihood of p_1

and n = the number of observations, and

k = the number of successes.

Two values, A and B are chosen such that if (1) $\lambda_n < A$, then hypothesis $H_1 : p = p_1$ is chosen, (2) $\lambda_n > B$, then hypothesis $H_2 : p = p_2$ is chosen, and (3) $A < \lambda_n < B$, then another item is chosen and another observation is made. Good approximations of A and B in terms of the two types of decision making error, α and β are specified as

(2) $A = \beta/(1-\alpha)$

$B = (1-\beta)/\alpha$

Combining equations (1) and (2) the inequality above may be written:

$$(3)\quad \frac{\beta}{(1-\alpha)} < \frac{p_2^k\,(1-p_2)^{n-k}}{p_1^k\,(1-p_1)^{n-k}} < \frac{(1-\beta)}{\alpha}$$

In order to simplify the computations, we take the logarithm of each term in (3) and rearrange terms so that (3) is expressed in terms of k, the number of successes.

$$(4)\quad \log\left(\frac{\beta}{1-\alpha}\right) < k \log\left(\frac{p_2}{p_1}\right) + (n-k)\log\left(\frac{1-p_2}{1-p_1}\right) < \log\left(\frac{1-\beta}{\alpha}\right)$$

$$(5)\quad \log\left(\frac{\beta}{1-\alpha}\right) < k \log\left(\frac{p_2}{p_1}\right) + n\log\left(\frac{1-p_2}{1-p_1}\right) - k\log\left(\frac{1-p_2}{1-p_1}\right) < \log\left(\frac{1-\beta}{\alpha}\right)$$

$$(6)\quad \log\left(\frac{\beta}{1-\alpha}\right) - n\log\left(\frac{1-p_2}{1-p_1}\right) < k\left[\log\left(\frac{p_2}{p_1}\right) - \log\left(\frac{1-p_2}{1-p_1}\right)\right] < \log\left(\frac{1-\beta}{\alpha}\right) - n\log\left(\frac{1-p_2}{1-p_1}\right)$$

$$(7)\quad \frac{\log\left(\frac{\beta}{1-\alpha}\right) - n\log\left(\frac{1-p_2}{1-p_1}\right)}{\log\left(\frac{p_2}{p_1}\right) - \log\left(\frac{1-p_2}{1-p_1}\right)} < k < \frac{\log\left(\frac{1-\beta}{\alpha}\right) - n\log\left(\frac{1-p_2}{1-p_1}\right)}{\log\left(\frac{p_2}{p_1}\right) - \log\left(\frac{1-p_2}{1-p_1}\right)}$$

Finally, we take advantage of the fact that the extremes of the inequality in (7) are linear functions of n, compute them, and graph them at the

beginning of the test. The value of k is then plotted as the test proceeds, and when it crosses one of the two straight lines, the test stops and a decision is arrived at.

Simple computational equations for the two straight lines are derived as follows:

$$(8)\quad d_1 = \frac{\log\left(\frac{\beta}{1-\alpha}\right) - n\log\left(\frac{1-p_2}{1-p_1}\right)}{\log\left(\frac{p_2}{p_1}\right) - \log\left(\frac{1-p_2}{1-p_1}\right)} \qquad \text{(left extreme (7))}$$

$$d_2 = \frac{\log\left(\frac{1-\beta}{\alpha}\right) - n\log\left(\frac{1-p_2}{1-p_1}\right)}{\log\left(\frac{p_2}{p_1}\right) - \log\left(\frac{1-p_2}{1-p_1}\right)} \qquad \text{(right extreme (7))}$$

$$(9)\quad g_1 = \log\left(\frac{p_2}{p_1}\right) \qquad g_2 = -\log\left(\frac{1-p_2}{1-p_1}\right)$$

$$a = \log\left(\frac{1-\beta}{\alpha}\right) \qquad b = -\log\left(\frac{\beta}{1-\alpha}\right)$$

$$h_1 = b/(g_1 + g_2) \qquad h_2 = a/(g_1 + g_2) \qquad s = g_2/(g_1 + g_2)$$

$$(10)\quad d_1 = -h_1 + sn \qquad \text{(lower line)}$$

$$d_2 = h_2 + sn \qquad \text{(upper line)}$$

In order to illustrate the operation of sequential testing strategies, an example will be carried out in detail. Assume that instruction which is effective for approximately 80% of the students is desired. High quality instruction will be defined as instruction which is effective for 90% of the students. p_1 is then 1.0-0.90 = 0.10. Low quality instruction will be defined as instruction which is effective for only 70% of the students.

p_2 is then 1.0-0.70 = 0.30. The instructional developer must then decide how much decision making error can be tolerated. For this example, assume that it is relatively costly to erroneously reject high quality instruction. Thus, a reasonable value for α, the probability of rejecting instruction that is effective for 90% of the target population, might be 0.01.

In many cases, the cost of erroneously accepting instruction that is of lower quality than that desired may be less than the cost of erroneously rejecting high quality instruction. The reasons for this lower cost vary from case to case, but revolve primarily around the fact that the instruction can always be improved if it is found to be unacceptable in practice. Under these circumstances a reasonable value for β, the probability of erroneously accepting instruction that is effective for 70% of the students, might be 0.10. The values for α and β refer directly to the values of p_1 and p_2.

The probability of rejecting instruction that is better than the instruction defined as high quality will always be less than α. Similarly, the probability of accepting instruction that is less effective than the instruction defined as low quality will always be less than β. In other words, α and β represent the decision making error that is tolerable for the worst possible case, if the instructional effectiveness is equal to p_1 or p_2. If the instruction is better or worse than the specified values of effectiveness, the errors in decision making will be less than α or β. Given these values, $p_1 = 0.10$, $p_2 = 0.30$, $\alpha = 0.01$, $\beta = 0.10$, it is possible to use the computational equations in (9) and (10) above to generate a sequential testing plan. The computations are shown in Table 2.

The first step in using the sequential plan is to plot lines d_1 and d_2

Necessary Parameters: $p_1 = .10$ $p_2 = .30$ $\alpha = .01$ $\beta = .10$

Calculations: $g_1 = \log\ p_2/p_1 = \log\ .30/.10 = \log\ 3 = .477$

$$g_2 = -\log \frac{1-p_2}{1-p_1} = -\log \frac{.70}{.90} = -\log (.7778) = .109$$

$$a = \log (1-\beta)/\alpha = \log\ .90/.01 = \log\ 90 = 1.954$$

$$b = -\log \frac{\beta}{1-\alpha} = -\log \frac{.10}{.99} = -\log (.1010) = .996$$

$$h_1 = b/(g_1 + g_2) = .996/(.477 + .109) = .996/.586 = 1.70$$

$$h_2 = a/(g_1 + g_2) = 1.954/(.477 + .109) = 1.954/.586 = 3.33$$

$$s = g_2/(g_1 + g_2) = .109/(.477 + .109) = .109/.586 = .186$$

$$d_1 = -h_1 + sn = -1.70 + 186n$$

$$d_2 = h_2 + sn = 3.33 + .186\ n$$

Table 2: Calculations for Sequential Plan to Evaluate Hypothetical Example

(Table 2). The instructional developer then administers the test of the objective to the first randomly chosen student. If he passes, a point one unit to the right of the zero point is plotted on the graph. If he fails, a point one unit to the right and one unit up from the zero point is plotted. The instructional developer then checks to see if he has crossed into either the acceptance region or the rejection region. If he has, the appropriate decision is made and further testing is unnecessary. If no decision can be made, another student is randomly selected, tested, and the same procedure is followed, starting at the previously plotted point rather than at zero. The procedure continues until a decision is reached. Table 3 contains hypothetical test results and subsequent actions by the instructional developer. Figure 1 shows the sequential plan based on the calculations in Table 2, which guided the decision making. The numbers which fall on the graph correspond to the student numbers in Table 3. The result of this hypothetical example is that the instruction is accepted with no immediate need for revision, based on a tryout sample of 15 students.

An alternative to the above plotting procedure is available, particularly if student test data can be gathered on a computer. The procedure simply calls for calculating the values of the extremes in equation (7). After each student attempts the test, the value of k, the total number of successes, is compared to the values of the extremes of the inequality. If k is less than the value of the left hand extreme, reject the instruction. If k is greater than the value of the right hand extreme, accept the instruction. If k falls between the extremes, continue sampling.

An obvious concern with sequential testing is: will the procedure ever terminate or will we forever remain in the region of indecision? Lindgren and McElrath (1966) state that Wald has shown that the test will terminate

Student Number	Test Results	Action	Decision
0		Begin at 0,0	Begin sampling
1	Pass	Move 1 unit right	No decision: continue sampling
2	Pass	Move 1 unit right	No decision: continue sampling
3	Pass	Move 1 unit right	No decision: continue sampling
4	Pass	Move 1 unit right	No decision: continue sampling
5	Pass	Move 1 unit right	No decision: continue sampling
6	Pass	Move 1 unit right	No decision: continue sampling
7	Pass	Move 1 unit right	No decision: continue sampling
8	Pass	Move 1 unit right	No decision: continue sampling
9	Fail	Move 1 unit right and 1 unit up	No decision: continue sampling
10	Pass	Move 1 unit right	No decision: continue sampling
11	Pass	Move 1 unit right	No decision: continue sampling
12	Pass	Move 1 unit right	No decision: continue sampling
13	Pass	Move 1 unit right	No decision: continue sampling
14	Pass	Move 1 unit right	No decision: continue sampling
15	Pass	Move 1 unit right	ACCEPT INSTRUCTION: STOP

Table 3: Hypothetical student test data used to implement a sequential testing plan

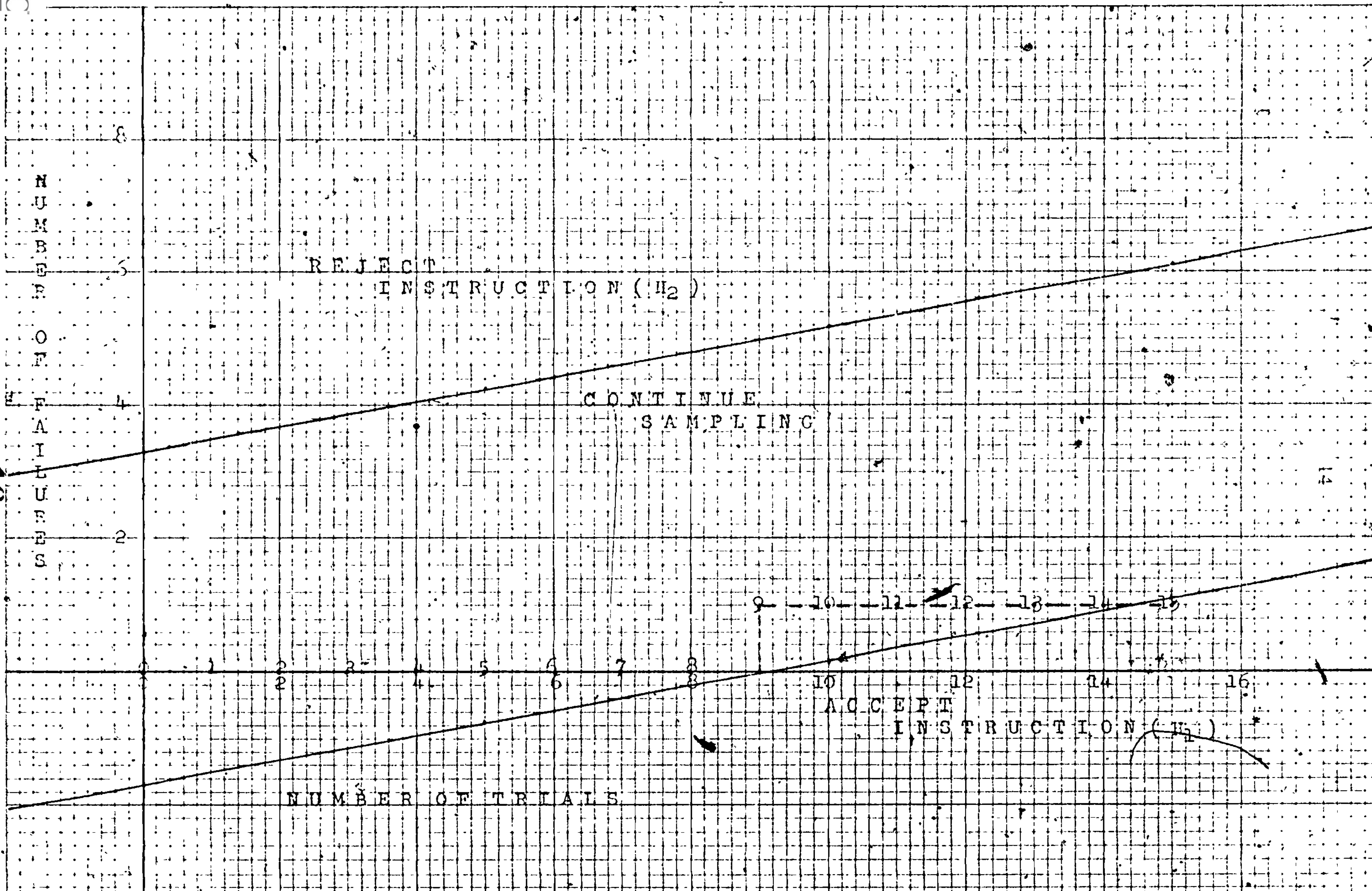

Figure 1: $p_1 = 0.10$ $p_2 = 0.30$ $\alpha = 0.01$ $\beta = 0.10$: Sequential Testing Plan

with probability one. Further, they point out that the number of observations required to reach a decision has an expected value that is usually less than the number of observations required to reach a decision with the same precision but using a fixed sample size. Crow, Davis and Maxfield (1960) discuss the use of truncated sampling plans to prevent the possibility of requiring a very large sample. Referring to equation (9), we agree to stop sampling when $n = (3ab)/(g_1g_2)$ (for the hypothetical example above this value equals approximately 112). This procedure results in negligible changes in the values of α and β. If n, the number of samples, gets this large with no decision, we accept the instruction, provided that the vertical distance from the n^{th} point to the lower line is less than the vertical distance from the n^{th} point to the upper line. Otherwise, we reject the instruction.

Baker and Alkin (1973) discuss the problems associated with empirically evaluating the usefulness of formative evaluation procedures. The sequential testing strategy suggested in this paper is as difficult to evaluate as other procedures. The major problem in judging the strength of any decision making procedure is the usual lack of suitable external criteria against which to compare decisions. However, some evidence that a sequential testing strategy is at least as useful as other procedures does exist. Two examples will be discussed.

The U. S. Army has been heavily involved in the design of audio-visual instruction to teach a wide variety of skills. Formative evaluation data was available for instruction to teach land navigation. The instruction covered eight objectives, each objective having associated with it a performance test that was scored pass/fail. Twenty-eight students participated in

the formative evaluation tryout. The tryout data is shown in Table 4.

Objective Number	Number Passing	Percent Passing	95% Confidence limits for proportion	
1	27	96	0.830	0.998
2	26	93	0.783	0.987
3	26	93	0.783	0.987
4	25	89	0.741	0.970
5	26	93	0.783	0.987
6	26	93	0.783	0.987
7	20	71	0.537	0.858
8	16	57	0.381	0.742
Overall*	26	93	0.783	0.987

* The overall criterion for passing was at least 6 of the 8 objectives accomplished.

Table 4: Results of U. S. Army Formative Evaluation for Audio-visual Instruction in Land Navigation, n = 28.

The decision rule used to evaluate this instruction was the 80/80 rule: 80% of the students pass 80% of the objectives. The 95% confidence limits imply that the instruction was certainly acceptable for objective #1, and that relatively high confidence can be placed in the effectiveness of the instruction for objectives #2,3,4,5, and 6. The effective for objective #7 is questionable, and the instruction for objective #8 is certainly below the minimum requirement. However, the overall data imply that the instruction may be accepted without further revision.

A sequential testing strategy was applied to the same data. The values for p_1, p_2, α, and β were the same values used in the hypothetical example discussed earlier. The results of the sequential testing procedure are summarized in Table 5. Figures 2a through 2h show the plotted data. Since there was no reason to believe that the students were arranged in any particular order, the results from student number 1 in the Army tryout were

plotted first, the results from student number 2 second, and so forth.

Objective Number	Number Tested	Decision
1	10	Accept
2	15	Accept
3	15	Accept
4	20	Accept
5	10	Accept
6	20	Accept
7	17	Reject
8	6	Reject

Table 5: Results of Sequential Testing for Formative Evaluation Data from Army Audio-visual Instruction in Land Navigation

Since I have argued that it is inappropriate to collapse across objectives, the overall data was not evaluated. The results of the sequential testing procedure agree with the results obtained using the 80/80 rule with 28 subjects; that is, we accept instruction for objectives #1 through #6, revise the instruction for objectives #7 and #8. In all cases, fewer students were needed than when using the 80/80 rule. In fact, the results for objective #8, clearly the objective for which revisions are most needed, were obtained with only 6 students.

Mitchell (1974) reported the results of using sequential testing for decision making during the development of instruction using interactive computer simulation. Four prototypes of the instruction were tried out and the results evaluated using a sequential testing strategy. Mitchell's values for the four necessary parameters were $p_1 = 0.80$, $p_2 = 0.50$, $\alpha = 0.01$, $\beta = 0.20$. Mitchell rejected the first three prototypes after only five students each attempted the instruction. The final version of the instruction

was accepted on the basis of data from four students. Other data collected as the students worked through the instruction itself, supported the rejection and acceptance decisions reached on the basis of the sequential testing. Mitchell's conclusions concerning the usefulness of the procedure are very encouraging:

"The sequential plans technique worked optimally for this validation. It did at the outset hold the promise of being more subject-efficient than earlier techniques and this was one of the reasons for its utilization (the other reason was the risk considerations included which are not normally found in the '80/80 criterion' or similar techniques)." (p. 25)

The major impetus behind this paper was dissatisfaction with commonly available decision making procedures for formative evaluation. A need exists for explicit information describing what is required of instructional de lopment efforts. It is also necessary to specify the risks that may be tolerated in accepting instruction as being ready for use or rejecting it and beginning the revision procedure. These requirements are particularly acute when the developer and the user are not the same people. Sequential testing strategies may help to make more explicit the formative evaluation process. At the same time, it may help to solve some of the resource demands of formative evaluation by decreasing the number of subjects required for the tryout sample. More experience will be required to learn how the procedure can best be implemented and to determine the value of the decisions made. Whatever the outcome of future research, the mere fact than an attempt is made to deal with formative evaluation in an objective manner will hopefully lead to improved instruction.

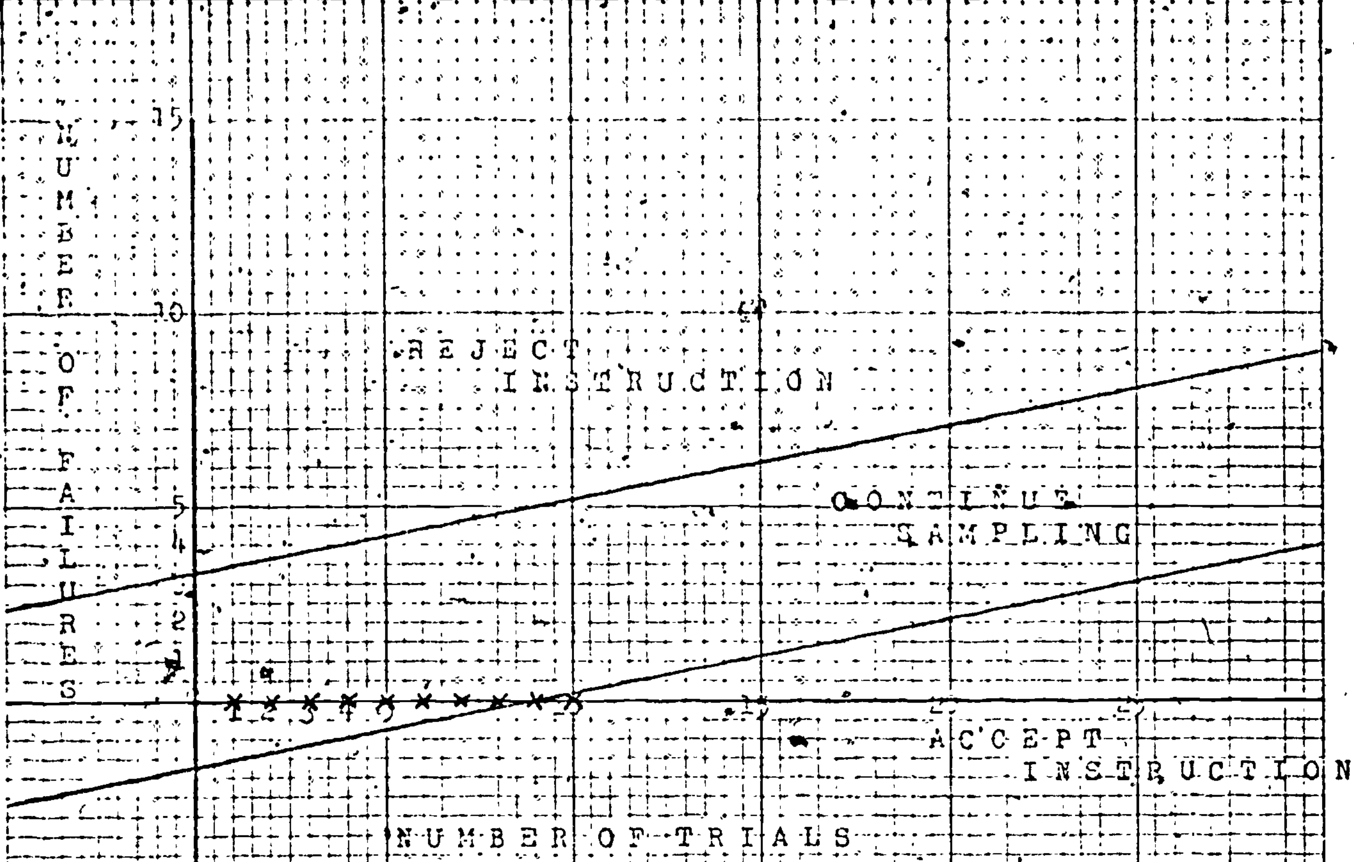

Figure 2a: Army audio-visual instruction Objective 1

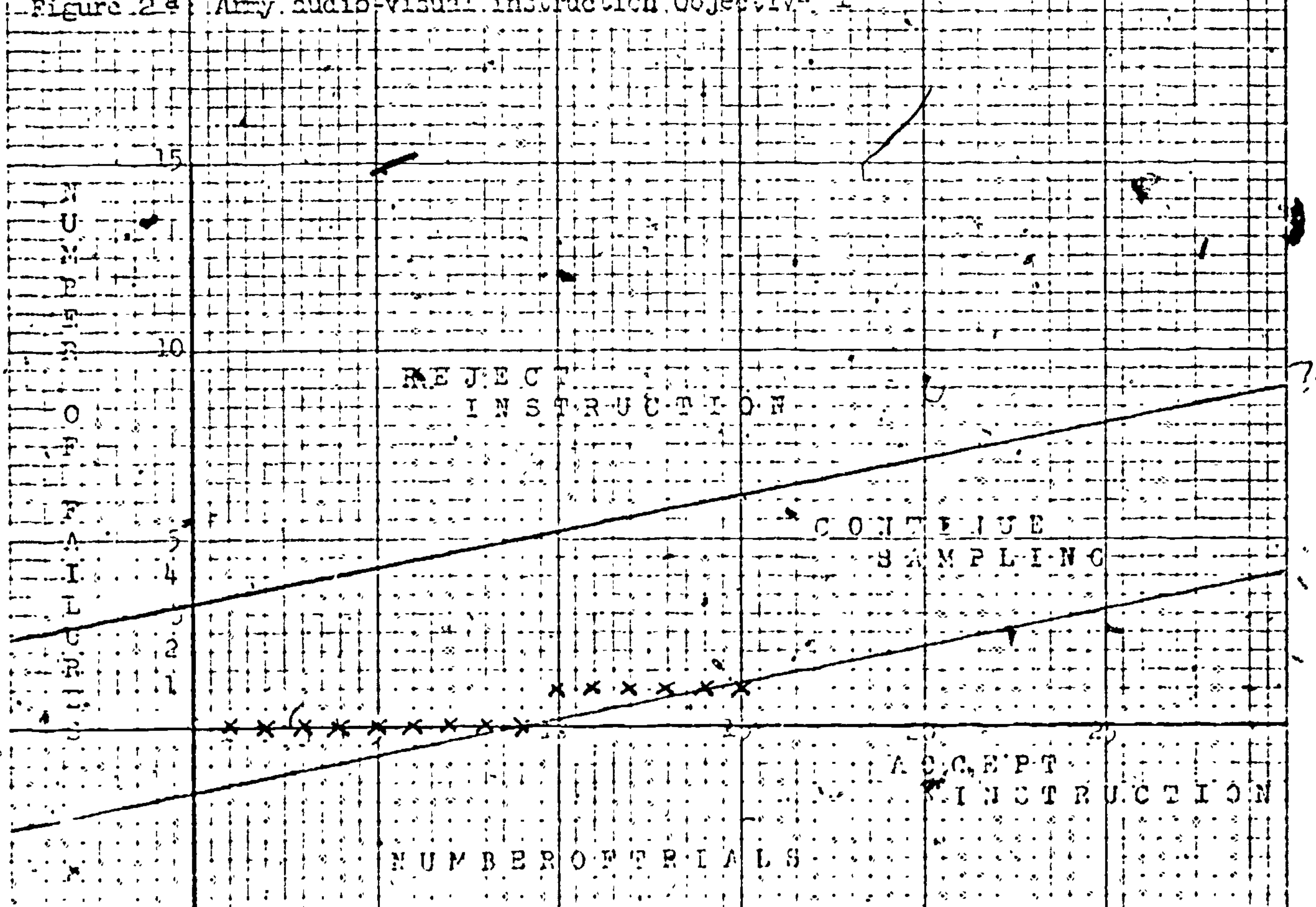

Figure 2b : Army audio-visual instruction Objective 2

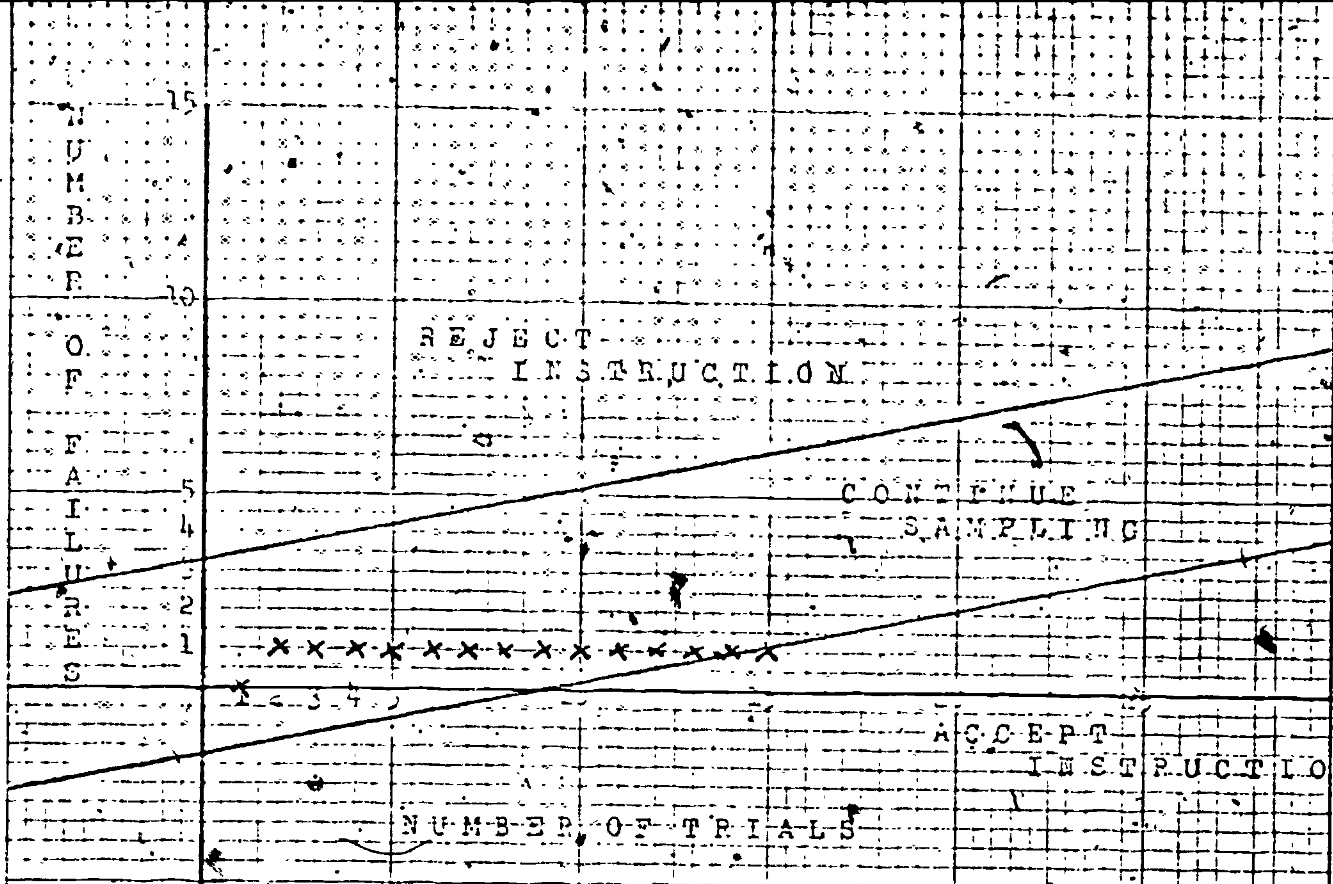

Figure 2c: Army audio-visual instruction Objective 3

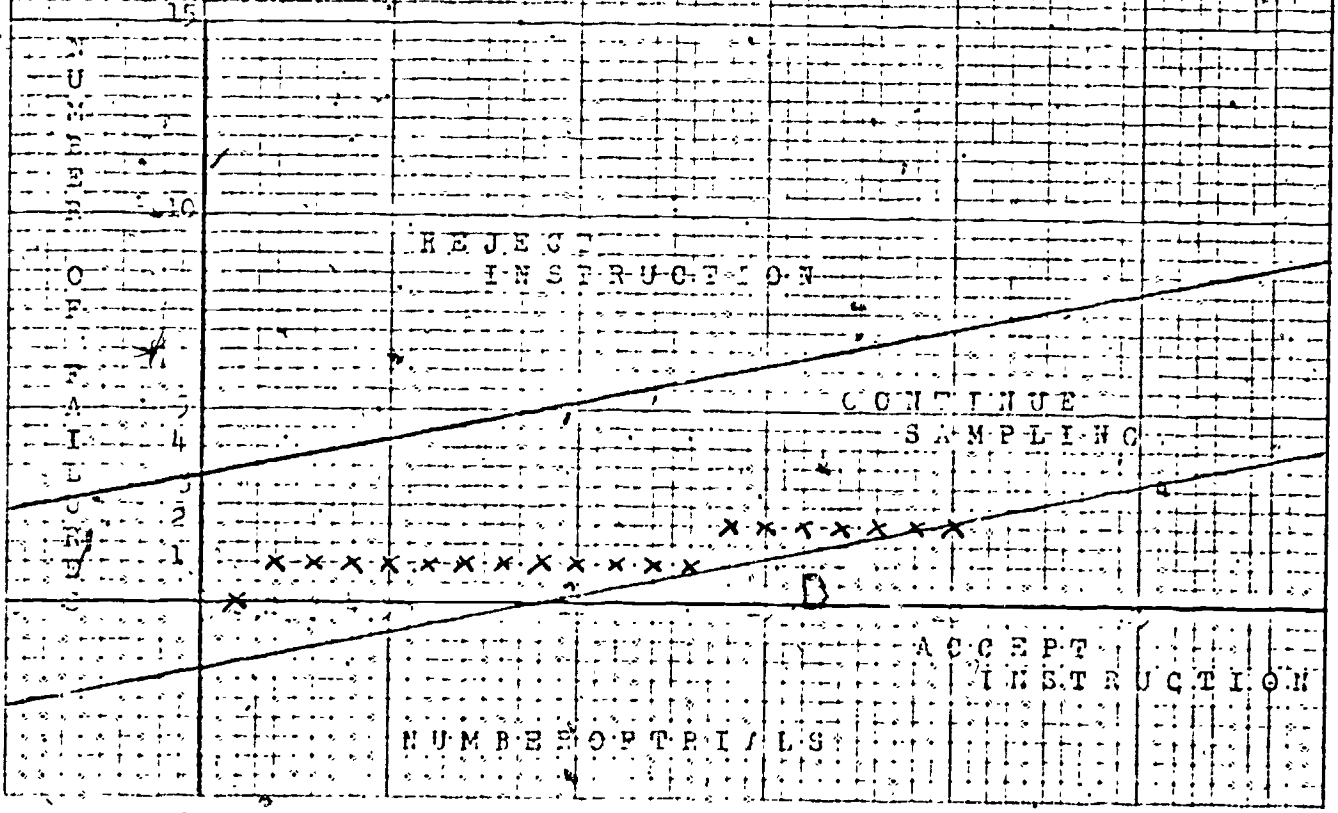

Figure 2d: Army audio-visual instruction Objective 4

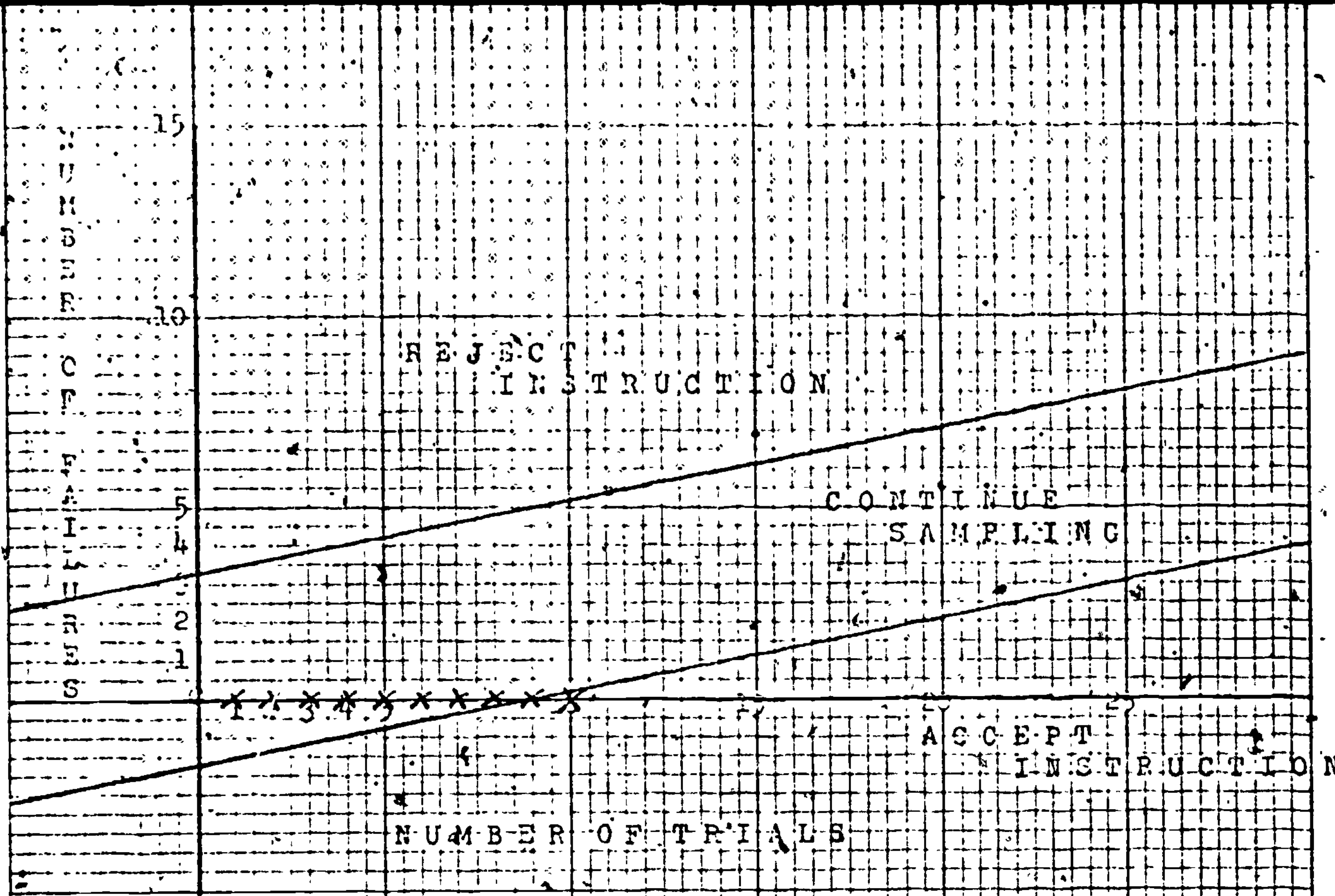

Figure 2 e: Army audio-visual instruction Objective 5

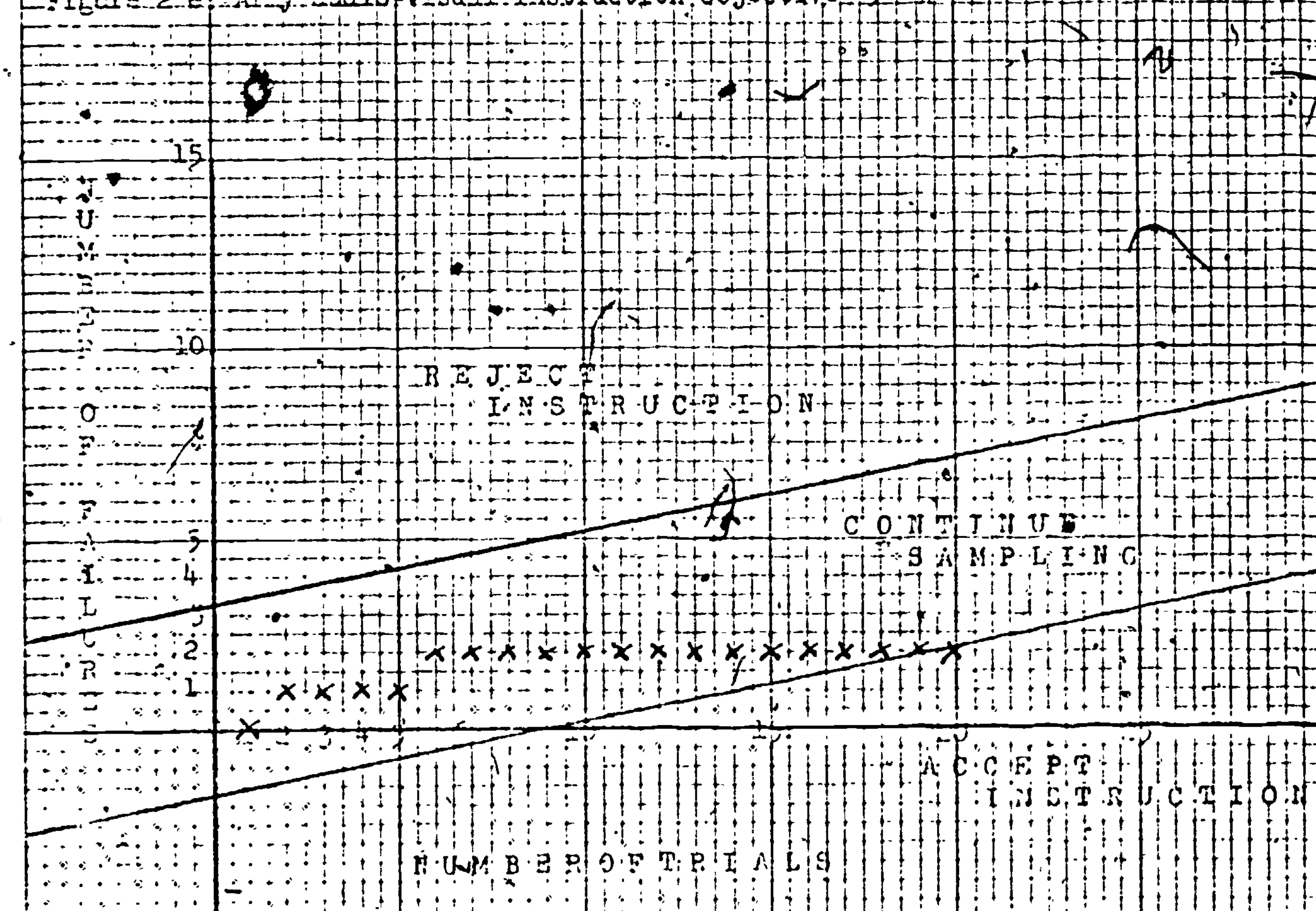

Figure 2 f : Army audio-visual instruction Objective 6

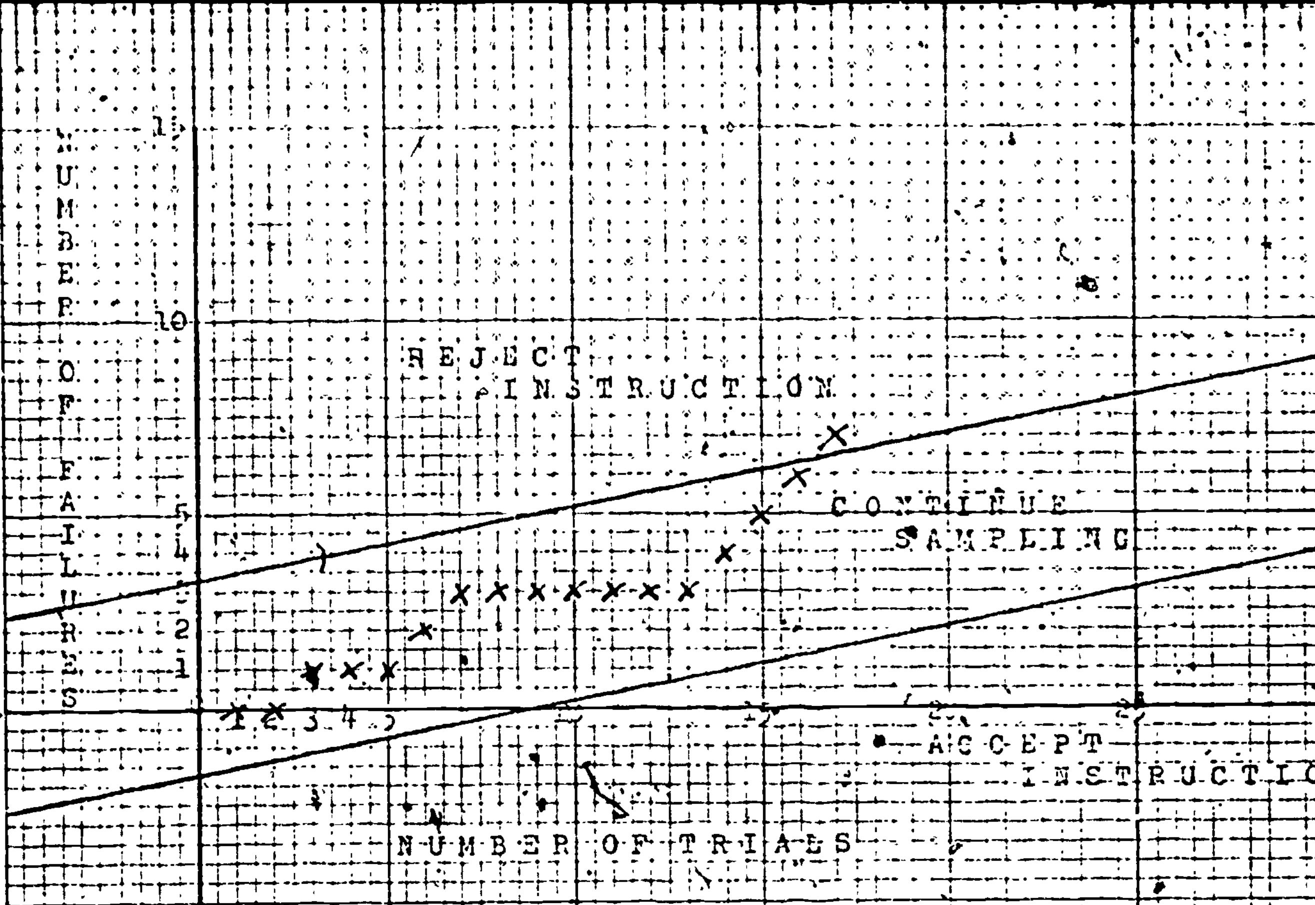

Figure 2 g: Army audio-visual instruction Objective 7

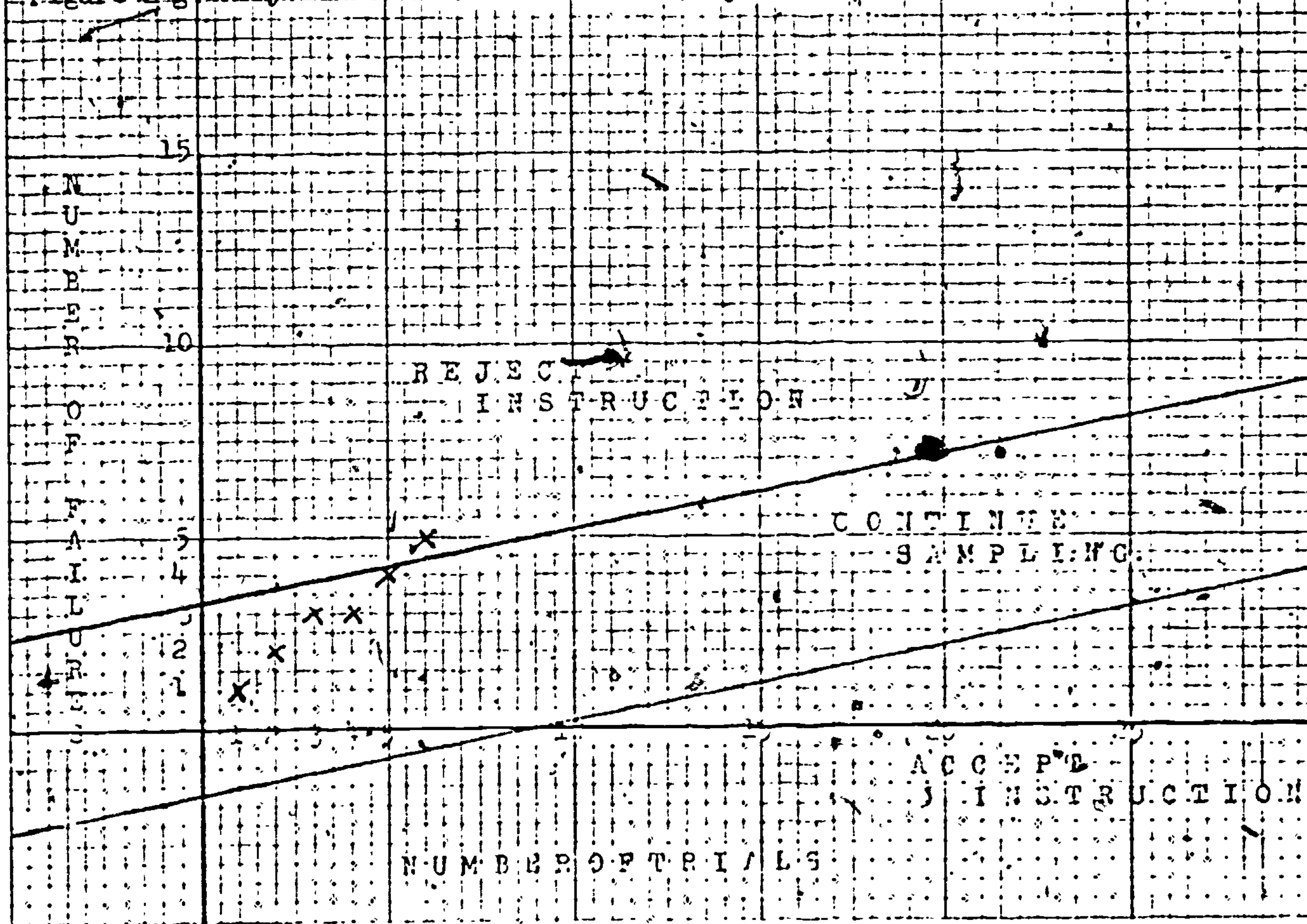

Figure 2 h: Army audio-visual instruction Objective 8

References

Baker, E. L. & Alkin, M. C. Annual review paper: Formative evaluation of instructional development. Audio-Visual Communication Review, 1973, 21, 389-418.

Crow, E. L., Davis, F. A , & Maxfield, M. W. Statistics Manual. New York: Dover Publications, Inc. 1960.

Lindgren, B. W. & McElrath, G. W. Introduction to Probability and Statistics. New York: The Macmillan Company, 1966.

Mitchell, M. C., Jr. The systematic design and validation of an interactive instructional computer simulation. Unpublished manuscript, Florida State University, 1974.